


HTML

HyperText Markup Language



```
<!DOCTYPE html>
<html>
<!-- created 2010-01-01 -->
<head>
<title>sample</title>
</head>
<body>
<p>Volutpatem accusantium
totam rem aperiam.</p>
</body>
</html>
```

HTML

Caractéristiques	
<u>Extensions</u>	.html, .htm, .xhtml, .xht
<u>Type MIME</u>	text/html, application/xhtml+xml
<u>PUID</u>	fmt/96
<u>Développé par</u>	World Wide Web Consortium & <u>WHATWG</u>
<u>Version initiale</u>	<u>1997</u>
<u>Type de format</u>	Langage de balisage
<u>Extension du</u>	<u>Standard Generalized</u> <u>Markup Language</u>
<u>Origine de</u>	<u>XHTML</u>
<u>Standard</u>	ISO/IEC 15445 <u>W3C HTML 4.01</u> <u>W3C HTML5</u>
<u>ISO</u>	15445
<u>Spécification</u>	<u>Format ouvert</u>
<u>Sites web</u>	<u>www.w3.org/html</u> <u>whatwg.org</u>

Dénominations

Évolution du langage

- 1989-1992 : Origine
- 1993 : Apports de NCSA Mosaic
- 1994 : Apports de Netscape Navigator
- 1995-1996 : HTML 2.0
- 1997 : HTML 3.2. et 4.0
- 2000 : XHTML
- 2007 à nos jours : HTML 5 et abandon du XHTML 2
- L'avenir du HTML : sans numéro de version ?

Description de HTML

- Syntaxe de HTML
- Structure des documents HTML
- Éléments de HTML
- Attributs de HTML
- Jeu de caractères

Interopérabilité de HTML

Notes et références

Voir aussi

Articles connexes

Liens externes

L'anglais *Hypertext Markup Language* se traduit littéralement en **langage de balisage d'hypertexte**¹. On utilise généralement le sigle HTML, parfois même en répétant le mot « langage » comme dans « langage HTML. *Hypertext* est parfois écrit *HyperText* pour marquer le *T* du sigle HTML.

Le public non averti parle parfois de HTM au lieu de HTML, HTM étant l'extension de nom de fichier tronquée à trois lettres, une limitation qu'on trouve sur d'anciens systèmes d'exploitation de Microsoft.

Durant la première moitié des années 1990, avant l'apparition des technologies web comme JavaScript, les feuilles de style en cascade et le *Document Object Model*, l'évolution de HTML a dicté l'évolution du *World Wide Web*. Depuis 1997 et HTML 4, l'évolution de HTML a fortement ralenti ; 10 ans plus tard, HTML 4 reste utilisé dans les pages web. En 2008, la spécification du HTML5 est à l'étude².

HTML est une des trois inventions à la base du World Wide Web, avec le Hypertext Transfer Protocol (HTTP) et les adresses web. HTML a été inventé pour permettre d'écrire des documents hypertextuels liant les différentes ressources d'Internet avec des hyperliens. Aujourd'hui, ces documents sont appelés «page web ». En août 1991, lorsque Tim Berners-Lee annonce publiquement le web sur Usenet, il ne cite que le langage SGML, mais donne l'URL d'un document de suffixe `.html`. Dans son livre *Weaving the web*³, Tim Berners-Lee décrit la décision de baser HTML sur SGML comme étant aussi « diplomatique » que technique : techniquement, il trouvait SGML trop complexe, mais il voulait attirer la communauté hypertexte qui considérait que SGML était le langage le plus prometteur pour standardiser le format des documents hypertexte. En outre, SGML était déjà utilisé par son employeur, l'Organisation européenne pour la recherche nucléaire (CERN). Les premiers éléments du langage HTML comprennent le titre du document, les hyperliens, la structuration du texte en titres, sous-titres, listes ou texte brut, et un mécanisme rudimentaire de recherche par index. La description de HTML est alors assez informelle et principalement définie par le support des divers navigateurs web contemporains. Dan Connolly a aidé à faire de HTML une véritable application de SGML⁴.

L'état de HTML correspond alors à ce que l'on pourrait appeler HTML 1.0. Il n'existe cependant aucune spécification portant ce nom, notamment parce que le langage était alors en pleine évolution. Un effort de normalisation était cependant en cours⁵. À partir de fin 1993, le terme HTML+ est utilisé pour désigner la version future de HTML⁶. Malgré l'effort de normalisation ainsi initié, et jusqu'à la fin des années 1990, HTML est principalement défini par les implémentations des navigateurs. Avec le navigateur NCSA Mosaic, HTML connaît deux inventions majeures. D'abord l'invention de l'élément IMG permet d'intégrer des images (dans un premier temps, uniquement aux formats GIF et XBM) aux pages web (Mosaic 0.10). Ensuite l'invention des formulaires (Mosaic 2.0pre5) rend le web interactif en permettant aux visiteurs de saisir des données dans les pages et de les envoyer au serveur web. Cette invention permet notamment de passer des commandes, donc d'utiliser le web pour faire du commerce électronique.

Avec l'apparition de Netscape Navigator 0.9 le 13 octobre 1994, le support de nombreux éléments de présentation est ajouté : attributs de texte, clignotement, centrage, etc. Le développement de HTML prend alors deux voies divergentes. D'une part, les développeurs de navigateurs s'attachent à maximiser l'impact visuel des pages web en réponse aux demandes des utilisateurs⁷. D'autre part, les concepteurs du web proposent d'étendre les capacités de description sémantique (logos, notes de bas de page, etc.) et les domaines d'applications (formules mathématiques, tables) de HTML. En ceci, ils suivent les principes de SGML consistant à laisser la présentation à un langage de style. En l'occurrence, les feuilles de style en cascade (CSS) sont prévues pour HTML. Seul le support des tables est rapidement intégré aux navigateurs, notamment parce qu'il permet une très nette amélioration de la présentation. Outre la multiplication des éléments de présentation, les logiciels d'alors produisant et consommant du HTML conçoivent souvent les documents comme une suite de commandes de formatage plutôt que comme un marquage représentant la structure en arbre aujourd'hui appelée *Document Object Model* (DOM). Le manque de structure du HTML alors mis en œuvre est parfois dénoncé comme étant de la « soupe balises », en anglais : *tag soup*.

1995-1996 : HTML 2.0

En mars 1995, le *World Wide Web Consortium* (W3C) nouvellement fondé propose le résultat de ses recherches sur HTML+ : le brouillon HTML 3.0. Il comprend notamment le support des tables, des figures et des expressions mathématiques. Ce brouillon expire le 28 septembre 1995 sans donner de suites directes. Fin 1995, le RFC 1866⁸ décrivant HTML 2.0 est finalisé. Le principal éditeur est Dan Connolly. Ce document décrit HTML tel qu'il existait avancé 1994, donc sans les nombreuses additions de Netscape Navigator

1997 : HTML 3.2. et 4.0

Le 14 janvier 1997, le W3C publie la spécification HTML 3.2. Elle décrit la pratique courante observée début 1996⁹, donc avec une partie des additions de Netscape Navigator et d'Internet Explorer. Ses plus importantes nouveautés sont la standardisation des tables et de nombreux éléments de présentation. HTML 3.2 précède de peu HTML 4.0 et contient des éléments en prévision du support des styles et des scripts. Le 18 décembre 1997, le W3C publie la spécification HTML 4.0 qui standardise de nombreuses extensions supportant les styles et les scripts, les cadres (*frames*) et les objets (inclusion généralisée de contenu). HTML 4.0 apporte également différentes améliorations pour l'accessibilité des contenus¹⁰ dont principalement la possibilité d'une séparation plus explicite entre structure et présentation du document, ou le support d'informations supplémentaires sur certains contenus complexes tels que les formulaires, les tableaux ou les sigles. HTML 4.0 introduit trois variantes du format, destinées à favoriser l'évolution vers un balisage plus signifiant, tout en tenant compte des limites temporaires des outils de production :

- la variante stricte (*strict*) exclut des éléments et attributs dits « de présentation », destinés à être remplacés par les styles CSS, ainsi que les éléments `applet` et `frame` qui sont remplacés par l'élément `object` conçu pour être plus apte à l'interopérabilité et à l'accessibilité.
- la variante transitoire (*transitional*) étend la variante stricte en reprenant les éléments et attributs dépréciés de HTML 3.2, dont les éléments de présentation sont couramment utilisés par les éditeurs HTML de l'époque.
- la variante *frameset* normalise la technique des jeux de cadres composant une ressource unique à partir de plusieurs pages web assemblées par le navigateur

Ces variantes perdurent par la suite sans modifications notables en HTML 4.01 et dans le format de transition XHTML 1.0 issu de HTML. La dernière spécification de HTML est la 4.01 datant du 24 décembre 1999. Elle n'apporte que des corrections mineures à la version 4.0.

2000 : XHTML

Le développement de HTML en tant qu'application du *Standard Generalized Markup Language* (SGML) est officiellement abandonné au profit de XHTML, application de *Extensible Markup Language* (XML). Cependant, en 2004, des fabricants de navigateurs web¹¹ créent le *web Hypertext Application Technology Working Group* (WHATWG) dans le but, notamment, de relancer le développement du format HTML et de répondre aux nouveaux besoins sur une base technologique jugée plus aisément implémentable que celle du XHTML 2.0 en cours de conception. Ceci s'inscrit dans le contexte d'une contestation plus générale du mode de fonctionnement du W3C, réputé trop fermé par une partie des développeurs et designers web¹².

2007 à nos jours : HTML 5 et abandon du XHTML 2

En mars 2007, tirant la conséquence des réticences d'une partie de l'industrie et des concepteurs de contenus web face à XHTML 2.0¹³, le W3C relance le développement de HTML et crée un nouveau groupe de travail encadré par Chris Wilson (Microsoft) et initialement Dan Connolly (W3C), maintenant Michael Smith (W3C). Il s'agit notamment¹⁴ :

- de faire évoluer HTML pour décrire la sémantique des documents mais aussi les applications en ligne ;
- de parvenir à un langage extensible via XML tout en maintenant une version non XML compatible avec les analyseurs syntaxiques existants (sers) HTML des navigateurs actuels ;
- et d'enrichir les interfaces utilisateurs avec des contrôles spécifiques : barres de progrès, menus, champs associés à des types de données spécifiques.

Les travaux du WHATWG ont été formellement adoptés en mai 2007 comme point de départ d'une nouvelle spécification HTML5¹⁵. Ce document¹⁶ a été publié sous forme de *Working Draft* le 22 janvier 2008. Parmi les principes de conception évoqués par le groupe de travail figurent en particulier¹⁷ :

- la compatibilité des futures implémentations HTML avec le contenu web existant, et la possibilité pour d'anciens agents utilisateurs d'exploiter les futurs contenus HTML 5 ;
- une approche pragmatique, préférant les évolutions aux modifications radicales, et adoptant les technologies ou pratiques déjà largement partagées par les auteurs de contenus actuels ;
- la priorité donnée, en cas de conflit d'intérêt, aux besoins des utilisateurs sur ceux des auteurs, et par suite, à ceux des auteurs sur les contraintes d'implémentation par les navigateurs ;
- le compromis entre la richesse sémantique du langage et l'utilité pratique des solutions disponibles pour remplir l'objectif majeur d'indépendance envers le média de restitution

Une *Accessibility Task Force* est créée par le W3C en novembre 2009 afin de résoudre les problèmes de compatibilité du nouveau format avec les normes d'accessibilité¹⁸, liés notamment à l'implémentation d'ARIA, aux alternatives textuelles et aux nouveaux éléments canvas et video¹⁹.

Le développement de XHTML 2.0 est initialement poursuivi en parallèle, en réponse aux besoins d'autres secteurs du web, tels que les périphériques mobiles, les applications d'entreprise et les applications serveurs²⁰. Puis, en juillet 2009, le W3C décide la non-reconduction du XHTML 2 Working Group à la fin 2009²¹.

Avec l'abandon du XHTML 2, la version XHTML 1.1 reste donc la version normalisée. Le HTML5 sera compatible avec le XHTML et le XML, et autorisera donc des documents XHTML5²². Cependant, il est probable que le W3C s'oriente vers un abandon pur et simple du XHTML 1.1, car l'implantation du XML dans le HTML5 rend inutile la définition de document de type XHTML.y.y (où y.y sont les numéros de version)²².

L'avenir du HTML : sans numéro de version ?

En janvier 2011, des divergences de points de vue entre Ian Hickson (ingénieur chez Google), qui écrit la spécification HTML5, et les membres du groupe de travail du W3C conduisent le WHATWG à créer *HTML Living Standard* (littéralement : standard vivant du HTML), une spécification de HTML prévue pour être en constante évolution, afin de coller avec les développements rapides de nouvelles fonctionnalités par les développeurs de navigateurs²³ (par opposition à des versions numérotées, donc « fixes »).

Le HTML Living Standard a pour but d'inclure le HTML5, et de le développer en permanence²⁴. En particulier, dans la version du 22 août 2012, le document de référence²⁴ explique que le HTML5 du W3C, publié le 22 juin 2012, est basé sur une version du HTML Living Standard, mais que le HTML Living Standard ne s'arrête pas à cette version, et continue à évoluer. Il développe en particulier les différences entre la version W3C (le HTML5) et la version HTML Living Standard (par exemple, les nouveaux bugs ne sont pas pris en compte dans le HTML5, des différences syntaxiques sont répertoriées, et de nouvelles balises créées par le HTML Living Standard ne sont pas incluses dans le HTML5).

Description de HTML

HTML est un langage de description de format de document qui se présente sous la forme d'un langage de balisage dont la syntaxe vient du *Standard Generalized Markup Language* (SGML).

Syntaxe de HTML

Jusqu'à sa version 4.01 comprise, HTML est formellement décrit comme une application du *Standard Generalized Markup Language* (SGML). Cependant, les spécifications successives admettent, par différents biais, que les agents utilisateurs ne sont pas, en pratique, des analyseurs SGML conformes²⁵. Les navigateurs Web n'ont jamais été capables de déchiffrer l'ensemble des variations de syntaxe permises par SGML²⁶ ; en revanche ils sont généralement capables de rattraper automatiquement de nombreuses erreurs de syntaxe, suivant la première partie de la « loi de Postel » : « Soyez libéral dans ce que vous acceptez, et conservateur dans ce que vous envoyez » (RFC 791²⁷). De fait, les développeurs de pages Web et de navigateurs Web ont toujours pris beaucoup de liberté avec les règles syntaxiques de SGML. Enfin, la *document type definition* (DTD) de HTML, soit la description technique formelle de HTML, n'a été écrite par Dan Connolly que quelques années après l'introduction de HTML⁴.

Malgré les libertés prises avec la norme, la terminologie propre à SGML est utilisée : document, élément, attribut, valeur, balise, entité, validité, application, etc. Grâce à la DTD, il est possible de vérifier automatiquement la validité d'un document HTML à l'aide d'un parseur SGML²⁸.

À l'origine, HTML a été conçu pour baliser (ou marquer) simplement le texte, notamment pour y ajouter des hyperliens. On utilisait un minimum de balises, comme dans le document HTML suivant :

```
<TITLE>Exemple de HTML </TITLE>
Ceci est une phrase avec un <A HREF=cible.html>hyperlien</A>.
<P>
Ceci est un paragraphe où il n'y a pas d'hyperlien.
```

Cet exemple contient du texte, cinq balises et une référence d'entité :

- <TITLE> est la balise ouvrante de l'élément TITLE.
- </TITLE> est la balise fermante de l'élément TITLE.
- Exemple de HTML est le contenu de l'élément TITLE.
- est la balise ouvrante de l'élément A, avec :
 - HREF=cible.html, l'attribut HREF dont la valeur est cible.html
- <P> est la balise ouvrante de l'élément P. Toutefois, elle est utilisée ici comme s'il s'agissait d'un séparateur de paragraphe, et c'est même ainsi qu'elle est souvent présentée dans les plus anciennes documentations de HTML. Il s'agit de la balise ouvrante du paragraphe dont le contenu est Ceci est un paragraphe où il n'y a pas d'hyperlien. La balise fermante de l'élément P, qui est optionnelle, est ici omise. L'élément P est implicitement terminé lorsqu'un nouveau paragraphe commence ou que l'élément parent est fermé (cas présent).
- ù est une référence d'entité représentant le caractère « ù ».
- Les balises peuvent être indifféremment écrites en minuscules ou majuscules. L'usage des minuscules devient plus courant car XHTML les impose.

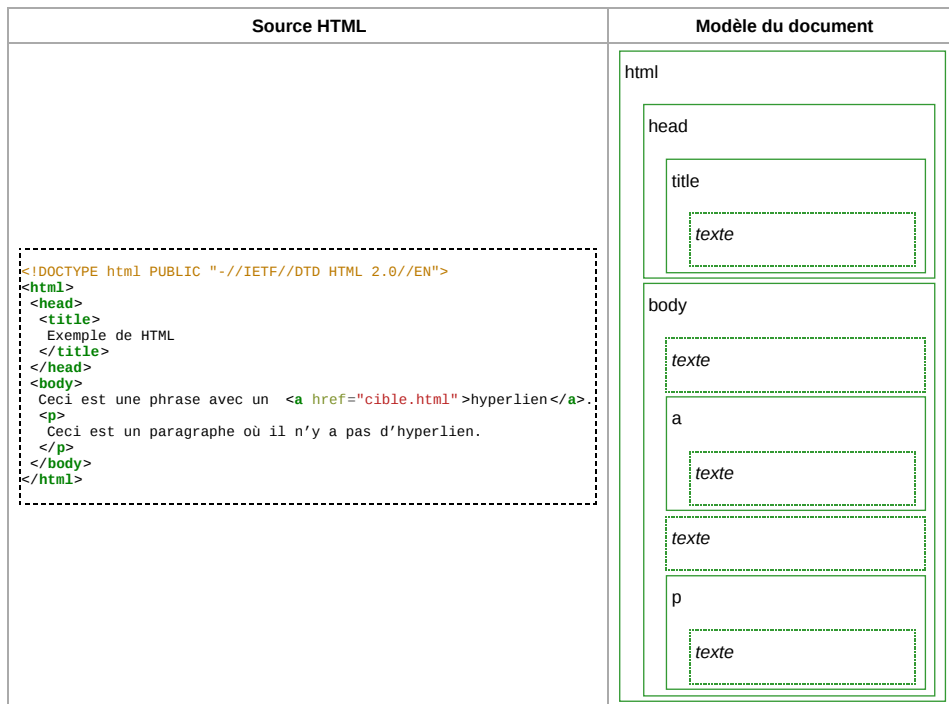
Un document HTML valide est un document qui respecte la syntaxe SGML, n'utilise que des éléments et attributs standardisés, et respecte l'imbrication des éléments décrite par le standard. Il ne manque qu'une déclaration de type de document à l'exemple précédent pour qu'il soit un document HTML 2.0 valide²⁹

Un document valide n'est cependant pas suffisant pour être conforme à la spécification HTML visée. En effet, outre l'exigence de validité, un document conforme est soumis à d'autres contraintes qui ne sont pas exprimées par la définition de type de document (DTD), mais qui le sont par la spécification elle-même. C'est notamment le cas du type de contenu de certains attributs, comme celui de l'attribut `datetime` : pour être conforme à HTML 4.01, celui-ci doit être lui-même conforme à un sous-ensemble de la norme ISO 8601³⁰. Un parseur strictement SGML tel que le validateur HTML du W3C ne peut donc pas garantir la conformité d'un document HTML.

Structure des documents HTML

Dans les premières années, les documents HTML étaient souvent considérés comme des structures plates, et les balises comme des commandes de style³¹. Ainsi la balise `<p>` était considérée comme un saut de ligne, et la balise `</p>` était ignorée. Ou encore lorsque JavaScript 1.0 est apparu, il ne donnait accès qu'aux liens et formulaires du document à travers les tables `document.forms` et `document.links`

Avec l'introduction des *Cascading Style Sheets* et du *Document Object Model*, il a fallu considérer que les documents HTML ont une véritable structure en arbre, avec un élément racine contenant tous les autres éléments³². Les balises ouvrantes et fermantes de ces éléments restent d'ailleurs optionnelles. Cependant, aujourd'hui, on a tendance à baliser chaque élément³³ et à indiquer la DTD. À l'exception de l'élément à la racine, chaque élément a exactement un élément parent direct ; cet « arbre du document » est notamment utilisé par la structure de formatage qui en est dérivée pour l'application des feuilles de style en cascade où chaque élément peut avoir un fond, un bord et une marge propres.



Éléments de HTML

La version 4 de HTML décrit 91 éléments. En suivant la spécification de HTML 4, les fonctionnalités implémentées par HTML peuvent être réparties ainsi :

Structure générale d'un document HTML³⁴

Au plus haut niveau, un document HTML est séparé entre un **en-tête** et un **corps**. L'en-tête contient les informations sur le document, notamment son titre et éventuellement des **métadonnées**. Le corps contient ce qui est affiché.

Informations sur la langue³⁵

Il est possible d'indiquer la langue de n'importe quelle partie du document et de gérer le **mélange de texte s'écrivant de gauche à droite avec du texte de droite à gauche**.

Marquage sémantique³⁶

HTML permet de différencier des contenus spécifiques tels que les citations d'œuvres externes, les extraits de code informatique, les passages en emphase et les abréviations. Certains de ces éléments, conçus initialement pour permettre le support de documentations techniques, sont très rarement employés (différenciation entre les éléments de *variable* et d'*exemple* de valeur dans un code informatique, par exemple, ou encore instance d'un terme défini dans le contexte).

Listes³⁷

HTML différencie des *listes non ordonnées* et des *listes ordonnées*, selon que l'ordre formel du contenu dans le code est en soi ou non une information. Des *listes de définition* existent également, mais sans que leur champ d'application ne soit exactement déterminé.

Tables³⁸

Cette fonctionnalité a été développée pour permettre la présentation de données tabulaires mais a été immédiatement exploitée pour ses puissantes capacités de mise en page.

Hyperliens³⁹

La fonctionnalité première de HTML.

Inclusion d'images, d'applets et d'objets divers⁴⁰

À l'origine HTML permettait seulement de donner des hyperliens sur les médias externes. L'invention d'éléments spécialisés pour le multimédia a permis l'inclusion automatique d'image, de musique, de vidéo, etc. dans les pages web.

Éléments de regroupement⁴¹

Ne conférant pas de signification au contenu qu'ils balisent, ces éléments génériques permettent d'appliquer des styles de présentation, de réaliser des traitements via des scripts ou tout autre opération nécessitant d'isoler une partie du contenu.

Style de la présentation⁴²

Chaque élément, voire tout le document, peut se voir appliquer des styles. Les styles sont définis dans le document ou proviennent de **feuilles de style en cascade (CSS)** externes.

Marquage de présentation du texte⁴³

Développé avant la généralisation de CSS pour fournir rapidement des fonctionnalités aux graphistes. D'usage désormais officiellement déconseillé pour la plus grande partie.

Cadres⁴⁴

Aussi connu sous le nom de *frames*, une fonctionnalité souvent décriée qui permet d'afficher plusieurs documents HTML dans une même fenêtre.

Formulaire pour l'insertion interactive de données⁴⁵

Les éléments de formulaire permettent aux visiteurs d'entrer du texte et des fichiers dans les pages Web.

Scripts⁴⁶

Permet d'associer des morceaux de programmes aux actions des utilisateurs sur le document. Les langages utilisés sont généralement JavaScript et VBScript.

Attributs de HTML

Les attributs permettent de préciser les propriétés des éléments HTML. Il y a 188 attributs dans la version 4 de HTML⁴⁷.

Certains attributs s'appliquent à presque tous les éléments :

- les attributs génériques `id` (identificateur unique) etc `class` (identificateur répétable)⁴⁸ destinés à permettre l'application de traitements externes, tels que l'application de styles de présentation ou de manipulation de l'arbre du document via un langage de script. Il s'y ajoute l'attribut `style`⁴⁹ permettant de définir le style de présentation de l'élément (généralement en CSS), et l'attribut `title`⁵⁰ apportant une information complémentaire de nature le plus souvent libre (l'exception majeure est l'utilisation de `title` pour déterminer le style permanent et les éventuels styles alternatifs appliqués à un document via des éléments `link`).

- les attributs d'internationalisation `dir` et `lang`³⁵ spécifiant la direction d'écriture et la langue du contenu ;
- les gestionnaires d'événement `onclick`, `ondblclick`, `onkeydown`, `onkeypress`, `onkeyup`, `onmousedown`, `onmousemove`, `onmouseout`, `onmouseover`, `onmouseup`⁵¹, qui capturent les évènements générés dans l'élément pour appeler un script.

D'autres attributs sont propres à un élément unique, ou des éléments similaires. Par exemple :

- les éléments qui permettent d'inclure dans le document des ressources graphiques sont dotés d'attributs de hauteur et de largeur afin que le navigateur puisse anticiper la taille de la ressource à afficher avant que celle-ci n'ait été téléchargée : `img`, `object`, `iframe`.
- des éléments spécifiques sont dotés d'un attribut assumant une fonction unique, tel que l'élément `table` des étiquettes des contrôles de formulaire et son attribut `for` désignant le contrôle concerné : c'est, en HTML, et avec les attributs `semap` et `ismap` des images, l'une des très rares associations explicites et formalisées entre des éléments, indépendamment de leur ordre linéaire dans le code source.

La plupart des attributs sont facultatifs. Quelques éléments ont cependant des attributs obligatoires :

- de par leur nature : l'élément `img` est obligatoirement doté d'un attribut `src` spécifiant l'URI de la ressource graphique qu'il représente. Il en est de même de tous les éléments dits « vides » et « remplacés »⁵² qui, au prix d'une entorse aux règles SGML, n'ont pas de contenu propre. C'est également le cas d'éléments non vides pour des raisons fonctionnelles, comme l'élément `form` dont l'attribut `action` indique la cible serveur qui traitera les données après soumission ;
- pour des raisons liées à l'accessibilité du contenu : les images sont ainsi dotées d'un attribut obligatoire `alt` permettant d'indiquer un contenu textuel brut destiné à remplacer la ressource graphique dans les contextes de consultation où elle ne peut pas être restituée ou perçue.

Le type de contenu des attributs HTML échappe pour partie au champ d'application de cette norme, et sa validation relève de normes tierces telles que les URI, les types de contenu ou les codes de langages.

Certains attributs sont enfin de type booléen. Ce sont les seuls attributs dont la syntaxe peut être valablement implicite en HTML : l'attribut `selected` d'un contrôle de formulaire peut ainsi être raccourci sous la forme `selected` remplaçant la forme complète `selected="selected"`. Cette forme particulière est un des points différenciant HTML de la syntaxe des documents « bien formés » au sens XML.

Jeu de caractères

Les pages Web peuvent être rédigées dans toutes sortes de langues et de très nombreux caractères peuvent être utilisés, ce qui requiert soit un jeu de caractères par type d'écriture, soit un jeu de caractères universel. Lors de l'apparition de HTML, le jeu de caractères universel Unicode n'était pas encore inventé, et de nombreux jeux de caractères se côtoyaient, notamment ISO-8859-1 pour l'alphabet latin et ouest-européen, Shift-JIS pour le japonais, KOI8-R pour le cyrillique. Aujourd'hui, le codage UTF-8 de Unicode est le plus répandu.

Le protocole de communication HTTP transmet le nom du jeu de caractères. En-tête HTML peut comporter le rappel de ce jeu de caractères, qui devrait être identique, sauf erreur de réglage. Enfin, à la suite d'un mauvais réglage, le jeu de caractères réellement utilisé peut encore différer du jeu annoncé. Ces mauvais réglages causent généralement des erreurs d'affichage du texte, notamment pour les caractères non couverts par la norme ASCII.

Avant la généralisation d'Unicode, des entités ont été définies pour représenter certains caractères non ASCII. Cela a commencé avec les caractères ISO 8859-1 dans la norme HTML 2.0. Pour les diacritiques, ces entités suivent un principe simple : la lettre suivie de l'abréviation de la diacritique associée.

Diacritiques d'ISO 8859-1

caractère	référence d'entité	remarque
Á	& Aacute ;	acute pour l'accent aigu
Â	& Acirc ;	circ pour l'accent circonflexe
Ã	& Agrave ;	grave pour l'accent grave
Ä	& Aring ;	ring pour le rond en chef
Å	& Atilde ;	tilde pour le tilde
Ä	& Auml ;	uml pour le tréma
Ç	& Ccedil ;	cedil pour la cédille
Ø	& Oslash ;	slash pour la barre oblique

Interopérabilité de HTML

Tel qu'il a été formalisé par le W3C, HTML est conçu pour optimiser l'interopérabilité des documents. Le HTML ne sert pas à décrire le rendu final des pages web. En particulier, contrairement à la publication assistée par ordinateur, HTML n'est pas conçu pour spécifier l'apparence visuelle des documents. HTML est plutôt conçu pour donner du sens aux différentes parties du texte : titre, liste, passage important, citation etc. Le langage HTML a été développé avec l'intuition que les appareils de toutes sortes seraient utilisés pour consulter le web : les ordinateurs personnels avec des écrans de résolution et de profondeur de couleurs variables, les téléphones portables, les appareils de synthèse et de reconnaissance de la parole, les ordinateurs avec une bande passante faible comme élevée, et ainsi de suite.

Comme HTML ne s'attache pas au rendu final du document, un même document HTML peut être consulté à l'aide de matériels et logiciels très divers. Au niveau matériel, un document peut notamment être affiché sur un écran d'ordinateur en mode graphique ou un terminal informatique en mode texte, il peut être imprimé, ou il peut être prononcé par synthèse vocale. Au niveau logiciel, HTML ne fait pas non plus de supposition, et plusieurs types de logiciels lisent le HTML : navigateur web, robot d'indexation, scripts divers (en Perl, PHP) de traitement automatique.

Un haut degré d'interopérabilité permet de baisser les coûts des fournisseurs de contenus car une seule version de chaque document sert des besoins très variés. Pour l'utilisateur du web, l'interopérabilité permet l'existence de nombreux navigateurs concurrents, tous capables de consulter l'ensemble du web.

Chaque version de HTML a essayé de refléter le plus grand consensus entre les acteurs de l'industrie, de sorte que les investissements consentis par les fournisseurs de contenus ne soient pas gaspillés et que leurs documents ne deviennent en peu de temps illisibles. La séparation du fond et de la forme n'a pas toujours été respectée au cours du développement du langage, comme en témoigne par exemple le balisage de style de texte, qui permet d'indiquer notamment la police de caractères souhaitée pour l'affichage, sa taille, ou sa couleur.

Notes et références

- (fr)** ISO/IEC 15445:2000, Langage de balisage d'hypertexte(http://www.iso.ch/iso/fr/CatalogueDetailPage.CatalogueDetail?CSNUMBER=27688&ICS1=35&ICS2=240&ICS3=30&scopelist=) (norme ISO).
- Le W3C publie le premier document de travail sur HTML 5, futur du contenu Web - La communauté Web établit le prochain standard HTML au sein du forum W3C public**(http://www.w3.org/2008/02/html5-pressrelease.html), communiqué du W3C daté du 22 janvier 2008.
- (en)** Tim Berners-Lee, Mark Fischetti, *Weaving the web: the past, present and future of the World Wide Web by its inventor*, Londres, Texere, 2000, p. 45-46 (ISBN 978-1-58799-018-2).
- (en)** James Gillies, Robert Cailliau, *How the Web was Born: The Story of the World Wide Web*, Oxford, Oxford University Press 2000, p. 212-213 (ISBN 978-0-19-286207-5).
- (en)** *Hypertext Markup Language (HTML)*(http://www.w3.org/MarkUp/draft-ietf-iiir-html-01.txt), *Internet Draft*, juin 1993.
- (en)** *HTML 4.01 Specification - 2.2.1 A brief history of HTML*(http://www.w3.org/TR/html4/intro/intro.html#h-2.2.1)- **(fr)** 2.2.1 Une brève histoire de HTML(http://www.la-grange.net/w3c/html4.01/intro/intro.html#h-2.2.1)

- Marc Andreessen, créateur de Netscape Navigatordéclarait dès 1993 sur la liste de discussion www-talk : « Je pense que s'occuper du SGML en général est une complète perte de temps, et que nous en serions aujourd'hui beaucoup plus loin si nous n'étions pas encombrés avec cet héritage SGML que nous continuons à porter 99,99 % des gens avec qui je parle veulent mettre en ligne des documents riches, veulent contrôler leur apparence, et se contre-fichent totalement du balisage sémantique ou des différences entre la structure et le rendu d'un document. »(en) HTML Spec (http://1997.webhistory.org/www.lists/www-talk.1993q2/0533.html) daté du 16 juin 1993.
- (en) Request for Commentsn^o 1866 (https://tools.ietf.org/html/rfc1866)
- (en) *HTML 3.2 Reference Specification, chap. Abstract* (http://www.w3.org/TR/REC-html32.html).
- La W3C a lancé en 1997 les premiers travaux de son (en) Initiative pour l'Accessibilité du web (http://www.w3.org/WAI/) (WAI), qui entame alors l'élaboration de ce qui allait devenir les premières(en) Directives pour l'Accessibilité des Contenus web (http://www.w3.org/TR/WCAG10/)(WCAG), essentiellement consacrées à l'accessibilité des documents HTML.
- L'initiative vient d'Opera et de Mozilla, rejointes par la suite par Apple, (en) http://www.w3.org/2004/04/webapps-cdf-ws/papers/opera.html
- (en) Jeffrey Zeldman,*An angry fix* (http://www.zeldman.com/2006/07/17/an-angry-fix/).
- Tim Berners Lee écrivait dèsoctobre 2006 : « la tentative de faire passer tout le monde à XML et aux contraintes syntaxiques que cela impose (guillemets autour des valeurs d'attributs, slashes dans les balises simples et espaces de noms), à intégrer d'un seul coup, n'a pas fonctionné … Il faut créer un groupe de travail HTML entièrement nouveau qui, contrairement au précédent, sera destiné à apporter des améliorations incrémentales à HTML, ainsi qu'à XHTML», (en) http://dig.csail.mit.edu/breadcrumbs/node/166
- Charte du groupe de travail HTML W3C.(en) http://www.w3.org/2007/03/HTML-WG-charter#deliverables
- (en) *Results of Questionnaire Shall we Adopt HTML5 as our specification text for review?* (http://www.w3.org/2002/09/wbs/40318/htmlg/results).
- (en) *HTML 5 – A vocabulary and associated APIs for HTML and XHTML – W3C Working Draft 22 January 2008* http://www.w3.org/TR/html5/.
- (en) *HTML Design Principles* (http://www.w3.org/html/wg/html5/principles).
- (en) *HTML Accessibility Task Force Work Statement, W3C* (http://www.w3.org/WAI/PF/html-task-force) *HTML Accessibility Task Force Wiki, W3C* (http://www.w3.org/WAI/PF/HTML/wiki/Main_Page)
- (en) Why is HTML5 Accessibility important? Because this child may never know HTML 4 (http://www.slideshare.net/stevefaulkner/html-5-accessibility), Steve Faulkner, juin 2009.
- « Le XHTML a été appliqué sur d'autres marchés, dont les périphériques mobiles, les applications d'entreprise, les applications serveurs, ainsi que dans un nombre croissant d'applications web telles que les blogiciels. Par exemple, le Groupe de travail des Meilleures Pratiques du web Mobile a inclus le standard XHTML Basic comme pierre angulaire des Meilleures Pratiques du web Mobile car les logiciels s'exécutant avec peu de mémoire peuvent le prendre en charge. … Avec la mise en place du Groupe de travail XHTML 2, leW3C continuera son travail technique sur le langage et prendra parallèlement en compte la revalorisation de la technologie afin d'expliquer clairement son indépendance et sa valeur sur le marché … »(en) Communiqué du W3C, 7 mars 2007 (http://www.w3.org/2007/03/html-pressrelease).
- (en) Communiqué du W3C, 2 juillet 2009 (http://www.w3.org/News/2009#item119)
- http://www.w3.org/2009/06/xhtml-faq.html
- http://www.clubic.com/internet/actualite-39258-html-standard-evolution-versions-specifique.html
- http://www.whatwg.org/specs/web-apps/current-work/multipage/introduction.html#is-this-html5?
- (en) Henri Sivonen,*An HTML5 Conformance Checker* (http://hsivonen.iki.fi/thesis/html5-conformance-checkerxhtml#history), *Master's Thesis*, Helsinki, 2007.
- Comme la forme courte<em/em> ou au lieu de emphaset.
- (en) Request for Commentsn^o 791 (https://tools.ietf.org/html/rfc791)
- Comme valideur HTML, on peut notamment citer le(en) *Markup Validation Service* (http://validator.w3.org/) du W3C.
- [1] (http://validator.w3.org/check?ss=1&doctype=HTML+2.0&fragment=%3Ctitle%3EExemple%20de%20HTML%3Ctitle%3E%0ACeci%20est%20une%20phrase%20avec%20un%20%3Ca%20href=cible.html%3Ehyperlien%3C/a%3E.%0A%3Cp%3E%0ACeci%20est%20un%20nouveau%20paragraphe%20où%20il%20n'y%20a%20pas%20d'hyperlien.)
- (en) *Marking document changes: The INS and DEL elements, HTML 4.01 Specification, W3C* (http://www.w3.org/TR/html401/struct/text.html#edef-datetime)
- Tim Berners-Lee considèrait les balises de **tit**e, et même de liste, comme des balises de style. Voir notamment la présentationLes standards HTML et CSS des origines à mercredi dernier (http://disruptive-innovations.com/zoo/slides/20060922/#semantic)de Daniel Glazman
- Des navigateurs commeNetscape Navigator4 ont eu du mal à faire cette transition ; ainsi un style appliqué à l'élémentp donnait des résultats dépendants de la présence de la balise optionnelle de fermeture</p>.
- Pour simplifier SGML, le standard XML (sur lequel est basé XHTML) n'autorise pas les balises optionnelles.
- (en) *HTML 4.01 Specification - 7 The global structure of an HTML document* (http://www.w3.org/TR/html401/struct/global.html)- (fr) 7 La structure globale du document HTML (http://www.la-grange.net/w3c/html4.01/struc/global.html)
- (en) *HTML 4.01 Specification - 8 Language information and text direction* (http://www.w3.org/TR/html401/struct/dirlang.html)- (fr) 8 Les indications de langue et la direction du texte (http://www.la-grange.net/w3c/html4.01/struc/dirlang.html)
- (en) *HTML 4.01 Specification - 9 Text* (http://www.w3.org/TR/html401/struct/text.html)- (fr) 9 Le texte (http://www.la-grange.net/w3c/html4.01/struc/text.html)
- (en) *HTML 4.01 Specification - 10 Lists* (http://www.w3.org/TR/html4/struct/lists.html)- (fr) 10 Les listes (http://www.la-grange.net/w3c/html4.01/struc/lists.html)
- (en) *HTML 4.01 Specification - 11 Tables* (http://www.w3.org/TR/html4/struct/tables.html)- (fr) 11 Les tables (http://www.la-grange.net/w3c/html4.01/struc/tables.html)
- (en) *HTML 4.01 Specification - 12 Links* (http://www.w3.org/TR/html401/struct/links.html)- (fr) 12 Les liens (http://www.la-grange.net/w3c/html4.01/struc/links.html)
- (en) *HTML 4.01 Specification - 13 Objects, Images, and Applets* (http://www.w3.org/TR/html401/struct/objects.html)- (fr) 13 Les objets, les images et les applets (http://www.la-grange.net/w3c/html4.01/struct/object.html)
- (en) *HTML 4.01 Specification - 7.5.4 Grouping elements: the DIV and SPAN elements* (http://www.w3.org/TR/html401/struct/global.html#edef-DIV)- (fr) 7.5.4 Le regroupement des éléments : les éléments DIV et SPAN (http://www.la-grange.net/w3c/html4.01/struct/global.html#edef-DIV)
- (en) *HTML 4.01 Specification - 14 Style Sheets* (http://www.w3.org/TR/html401/present/styles.html)- (fr) 14 Les feuilles de style (http://www.la-grange.net/w3c/html4.01/present/styles.html)
- (en) *HTML 4.01 Specification - 15 Alignment, font styles, and horizontal rule* (http://www.w3.org/TR/html401/present/graphics.html)- (fr) 15 L'alignement, les styles de police et les règles horizontales (http://www.la-grange.net/w3c/html4.01/present/graphics.html)
- (en) *HTML 4.01 Specification - 16 Frames* (http://www.w3.org/TR/html401/present/frames.html)- (fr) 16 Les cadres (http://www.la-grange.net/w3c/html4.01/present/frame.html)
- (en) *HTML 4.01 Specification - 17 Forms* (http://www.w3.org/TR/html401/interact/forms.html)- (fr) 17 Les formulaires (http://www.la-grange.net/w3c/html4.01/interact/forms.html)
- (en) *HTML 4.01 Specification - 18 Scripts* (http://www.w3.org/TR/html4/interact/script.html)- (fr) 18 Les scripts (http://www.la-grange.net/w3c/html4.01/interact/scripts.html)
- (en) *HTML 4.01 Specification - Index of Attributes* (http://www.w3.org/TR/html401/index/attributes.html)- (fr) Index des attributs (http://www.la-grange.net/w3c/html4.01/index/attributes.html)
- (en) *HTML 4.01 Specification - 7.5.2 Element identifiers: the id and class attributes* (http://www.w3.org/TR/html401/struct/global.html#h-7.5.2)- (fr) 7.5.2 Les identifiants des éléments : les attributs id et class (http://www.la-grange.net/w3c/html4.01/struct/global.html#h-7.5.2)
- (en) *HTML 4.01 Specification - 14.2.2 Inline style information* (http://www.w3.org/TR/html401/present/styles.html#h-14.2.2)- (fr) 14.2.2 Les informations de style en-ligne (http://www.la-grange.net/w3c/html4.01/present/styles.html#h-14.2.2)
- (en) *HTML 4.01 Specification - 7.4.3 The title attribute* (http://www.w3.org/TR/html401/struct/global.html#h-7.4.3)- (fr) 7.4.3 L'attribut title (http://www.la-grange.net/w3c/html4.01/struct/global.html#h-7.4.3)
- (en) *HTML 4.01 Specification - 18.2.3 Intrinsic events* (http://www.w3.org/TR/html401/interact/scripts.html#h-18.2.3)- (fr) 18.2.3 Les événements intrinsèques (http://www.la-grange.net/w3c/html4.01/interact/scripts.html#h-18.2.3)
- au sens de CSS 2.1 qui explicite la notion d'« élément remplacé » restée implicite en HTML. Voir (en) http://www.w3.org/TR/CSS21/conform.html#replaced-element

Voir aussi

Articles connexes

- World Wide Web
- World Wide Web Consortium (W3C)
- Élément HTML
- Identificateur de fragment
- Extensible Hypertext Markup Language(XHTML)
- HTML5
- SGML
- SELFHTML une documentation en ligne de la syntaxe HTML
- JHTML

Liens externes

- (en) The W3c Markup Validation Service

Documents officiels :

Sur les autres projets Wikimedia :



HTML, sur Wikimedia Commons



HTML, sur le Wiktionnaire



Hypertext Markup Language sur Wikiversity



Hypertext Markup Language sur Wikibooks

- **(en)** « [RFC1866 : Hypertext Markup Language - 2.0](#) », novembre 1995
 - **(en)** [HTML 3.2 Reference Specification](#) recommandation du W3C, 14 janvier 1997
 - **(en)** [HTML 4.0 Specification](#) recommandation du W3C, 18 décembre 1997
 - **(en)** [HTML 4.01 Specification](#) dernière recommandation HTML 4 du W3C, 24 décembre 1999
 - **(en)** [HTML 5 A vocabulary and associated APIs for HTML and XHTML](#) Lactuel brouillon de travail du W3C sur HTML5, 25 août 2009
 - **(en)** [HTML 5 differences from HTML 4](#) brouillon de travail du W3C sur les différences entre HTML4 et HTML5, 25 août 2009
 - **(en)** [HTML Living Standard specifications](#) la documentation du HTML Living Standard, successeur du HTML5, qui n'a plus de version numéroté.
 - **(fr)** [La spécification HTML 4.01](#) traduction non normative
-

Ce document provient de «https://fr.wikipedia.org/w/index.php?title=Hypertext_Markup_Language&oldid=147407274».

La dernière modification de cette page a été faite le 10 avril 2018 à 22:26.

Droit d'auteur : les textes sont disponibles sous licence Creative Commons attribution, partage dans les mêmes conditions ; d'autres conditions peuvent s'appliquer. Voyez les conditions d'utilisation pour plus de détails, ainsi que les crédits graphiques. En cas de réutilisation des textes de cette page, voyez comment citer les auteurs et mentionner la licence. Wikipedia® est une marque déposée de la Wikimedia Foundation, Inc., organisation de bienfaisance régie par le paragraphe 501(c)(3) du code fiscal des États-Unis.